

Петрасова С.В.

Национальный технический университет

«Харьковский политехнический институт»

Научный руководитель – доцент Хайрова Н.Ф.

ФОРМАЛИЗАЦИЯ СЕМАНТИЧЕСКИХ ОТНОШЕНИЙ КОНЦЕПТОВ ТОЛКОВОГО СЛОВАРЯ

В настоящее время стремительный рост объемов информации приводит к необходимости создания и использования интеллектуальных информационных систем, с помощью которых возможно систематизировать и обрабатывать накопленные знания.

Семантическая сеть сегодня является наиболее перспективным способом формального выражения знаний. Это обусловлено, прежде всего, наглядностью представления знаний в семантических сетях. Кроме того, с их помощью возможно в явном виде выражать семантические отношения между словами [1].

Проведенный анализ показал, что область применения семантических сетей достаточно обширная. Семантические сети используются для расширения электронных тезаурусов; как лингвистическое обеспечение в информационно-поисковых системах; в системах обработки естественного языка, экспертных системах и т.д.

Но, несмотря на достаточно широкое использование данного способа представления знаний, распространение семантических сетей сдерживается как неоднозначностью представления естественного языка, так и трудоемкостью и сложностью их разработки.

В своем подавляющем большинстве семантические сети разрабатываются экспертами в той или иной предметной области «вручную». Тогда как одним из наиболее полных источников знаний являются такие универсальные средства представления, накопления и передачи информации в человеческом обществе, как тексты. Среди всех полнотекстовых источников толковые словари представляют, пожалуй, тексты с наиболее концентрированной смысловой нагрузкой. Так как именно в данном информационном ресурсе каждая словарная статья содержит значимые слова, отражающие конкретную предметную область.

В данном исследовании предлагается метод автоматизированного построения фрагмента семантической сети на базе знаний толкового словаря английского языка. В разработанной сети используются такие отношения, как принадлежности классу, гипонимия, меронимия, гиперонимия и отношение семантической эквивалентности.

Под семантическими эквивалентами мы понимаем текстовые выражения, сопоставленные одному и тому же понятию, а также слова и словосочетания с близким значением, встречающиеся в одном контексте. Для формализации отношений семантической эквивалентности определяется мера семантической близости, формально определяемая дефинициями глоссариев как отношение мощностей множеств, образованных теоретико-множественным пересечением и объединением множеств терминов дефиниций [2]. Преимуществом данного метода является комбинация статистической и семантической составляющих, что позволяет дать количественную оценку семантической близости между словами.

В разработанной семантической сети для формализации межконцептуальных отношений принадлежности классу, гипонимии, меронимии и гиперонимии используются шаблоны лексических последовательностей. Например, формальным выражением отношения принадлежности классу может быть шаблон $NN_{MoC} - Rel - NN_{Class}$, где MoC – член класса, Class – класс, Rel – лексические цепочки, выражающие отношение классификации, например, *meaning of, identification of, is a* и др.

В результате исследования разработан и программно реализован алгоритм автоматизированного построения семантической сети, выявляющий вышеприведенные отношения между терминами толкового словаря.

Для определения эффективности разработанного информационно-лингвистического обеспечения исследовалась выборка из ста терминов словаря Microsoft Computer Dictionary [3]. После обработки полученных результатов были определены коэффициенты точности, полноты и ошибки.

Разработанное программное обеспечение может быть использовано для расширения существующих электронных тезаурусов; в информационно-поисковых системах; в системах автоматизированной обработки естественного языка, в частности, для решения проблемы синонимии.